

# Adaptive Recommendation and Open-Ended Semiosis

LUIS MATEUS ROCHA

*Complex Systems Modeling Team*

*Computer Research and Applications Group*

*Los Alamos National Laboratory, MS B265*

*Los Alamos, NM 87545, U.S.A.*

*E-mail: [rocha@lanl.gov](mailto:rocha@lanl.gov) WWW: <http://www.c3.lanl.gov/~rocha>*

*International Journal of Human-Computer Studies. (In Press)*

**Abstract:** A recommendation system for an extended process of information retrieval in distributed information systems is proposed. This system is both a model of dynamic cognitive categorization processes and powerful real application useful for knowledge management. It utilizes an extension of fuzzy sets named evidence sets as the mathematical mechanisms to implement the categorization processes. It is a development of some aspects of Pask's Conversation Theory. It is also an instance of the notion of linguistic-based selected self-organization here described, and as such it instantiates an open-ended semiosis between distributed information systems and the communities of users they interact with. This means that the knowledge stored in distributed information resources adapts to the evolving semantic expectations of their users as these select the information they desire in conversation with the information resources. This way, this recommendation system establishes a mechanism for user-driven knowledge self-organization.

**Keywords:** Recommendation systems, information retrieval, knowledge management, fuzzy sets, evidence sets, distributed information systems, evolutionary systems, cognitive categorization, self-organization, selection, human-machine interaction, artificial intelligence, semiotics, cognitive science, Gordon Pask.

## INTRODUCTION: PASK AND EVOLVING HUMAN-MACHINE INTERACTION

Gordon Pask sought underlying principles of organization and communication which dealt with the necessity of incorporating the subjectivity of human experience. Gordon's primary contribution to cybernetics and systems science was his emphasis on the personal nature of reality, and on the process of learning as stemming from the consensual agreement of interacting actors in a given environment. Life and intelligence lie somewhere in the conflict between closed, unique, construction and open, shared, interaction. Between a specific material fabric, and a social conceptual organization.

He developed an extensive theory of conversation [Pask, 1975, 1976] that proposed the abandonment of the concept of learning as a one to one mapping of real world to mental categories, for a dynamic, internal, self-organizing process of coming to know, constrained by developmental interaction with an environment and fellow "knowers". It is interesting to notice that his constructivist message of a personal reality lead to a research program particularly interested in devising ways of improving human-human and human-machine

interaction as a means of developing consensual scientific progress. He was interested in artificial systems as vehicles for driving through knowledge in a new level of human-machine symbiosis aiming at increasing world-wide understanding.

The work here presented is both an attempt to re-conceptualize Gordon's ideas in a more current terminology and to develop a computer system that follows in essence some of his goals. This link is not fully explained until the end of the present work, but Gordon's influence is undoubtedly felt throughout. In section 1, the idea of language as a mechanism to achieve an open-ended semiosis between cognitive systems and their environments is proposed. In section 2, the process of cognitive categorization is discussed and evidence sets are presented as candidates for the modeling of cognitive categories. In section 3 an adaptive recommendation system for information retrieval, *TalkMine* is presented which can organize knowledge according to the evolving needs of its users. *TalkMine* is both a model of cognitive categorization as a means to recombine self-organize knowledge and is also a powerful system which solves some of the shortcomings of current recommendation systems. discussed in sections 1 and 2. Finally, in section 4, this system is shown to be an extension of Pask's ideas which implements an evolving, open-ended semiosis between communities of users and distributed information resources.

## 1. SELECTED SELF-ORGANIZATION AND OPEN-ENDED SEMIOSIS

### 1.1 Selected Self-Organization: From Classification to Categorization

We know a good deal from ethology, enough to realize that one of the prevalent structuring processes of the animal brain is the propensity to deconstruct observable objects or events and then to respond based upon an assessment of the elemental parts. As Lorenz [1971] has shown, a goose will 'see' not an egg as it will see elements of an egg such as color, speckled pattern, shape, and size. A goose can be easily fooled into sitting on a nest of wooden eggs with these elements exaggerated (a brighter green, a more perfect ovoid, larger speckles, and the like). The deconstruction of reality into elements grants obvious and powerful survival potential: an animal will not be focused on one specific egg, but can recognize all eggs. Events and situations can have different components but elicit the same survival responses. Categorical flexibility is a desired trait in varying environments.

The self-organizing or connectionist paradigm in systems research and cognitive science, has rightly emphasized these characteristics of mental behavior. A given dynamics, say the neuronal interactions of the brain or a cellular chemical network, will converge to a number of attractor states. This is often referred to as a process of self-organization. Such a dynamic self-organizing system may then utilize these attractors to classify its own interactions with an environment. The ability of a self-organizing system to relate internal stabilities to aspects of its environmental coupling has been referred to as *emergent classification* elsewhere [Rocha, 1996, 1997a]. Clearly, self-organizing systems, if not chaotic, will classify similar events in their environments to similar attractor points of their dynamics: the categorical flexibility observed above.

However, to effectively deal with a changing environment, systems capable of relating internal stabilities to environmental regularities, must be able to change their own dynamics in order to create new basins of attraction for new classifications. In other words, the self-organizing system must be structurally perturbed by or coupled [Maturana and Varela, 1987] to some external system which acts on the structure of the first inducing some form of explicit or implicit selection of its dynamic classifications, this has been referred to as *selected self-organization* in [Rocha, 1996, 1997a, 1998a]. Now, for selection to occur we must have some internal vehicle for classification — there must be different alternatives. The attractor landscape of self-

organizing systems offers these alternatives. One way of conceptualizing this is to think of the attractor landscape as a distributed memory bank [van Gelder, 1992], where each attractor basin is seen as storing a given classification configuration. I refer to this ability of a self-organizing system to select appropriate dynamic classification configurations to deal with a changing environment as *categorization* (or *semantic emergence* in the context of evolutionary systems [Rocha, 1996, 1998a]). A category is then a higher level grouping of dynamic classifications related in some environmental context.

In the biological realm, such selection is implicitly defined by different rates of reproduction of individuals in varying (genetic) populations, while in the cognitive realm we have some form of more explicit selection based on learning and cultural evolution<sup>1</sup>. A simple example in an applied domain, would be an external algorithm for selecting the weights of a neural network, or some other connectionist device, in order to achieve some desired classification.

## 1.2 Improving Selected Self-Organization: Linguistic Recombination

A relevant question at this point is how effective can this selected self-organization get? Connectionist machines can only classify current inputs, that is, they cannot manipulate their own distributed records. Structural perturbation can alter their classification landscape, but we do not have a process to actually access a particular category at any arbitrary time, except by re-presenting the inputs that cause it to the network. Something similar happens in the biological realm. Biological reproduction is a process of phenotypical construction from instructions stored in memory as conceptually defined by von Neumann [1966; see Pattee, 1995; Rocha, 1996]. If living systems were purely self-organizing systems subjected to selection, then reproduction would have to rely on components that could replicate themselves in a template fashion, or components that could unfold and fold at will so that copies could be made from available elements. In other words, if living systems did not have a symbolic dimension in DNA, life forms would be restricted to those proteins and enzymes that could reproduce in a crystal-like manner, or that could unfold to be reconstructed from available amino acids, and then re-fold to their original form.

Indeed, DNA introduces a novel dimension to living systems which allows them to construct any protein from a genetic description, and not only those that can self-reproduce in the above described senses. This way, DNA introduces a kind of random access memory so that living systems have access at any time to the blueprints of their own construction. This ability liberates living systems from purely localized interactions; biological reproduction is not restricted to template reproduction as the genetic, localized, descriptions can be communicated much more effectively from generation to generation, as well as to different parts of organisms. This kind of selected self-organization can be referred to as *description-based selected self-organization* [Rocha, 1996, 1998a].

In biological systems the existence of a (genetic) code or syntax allows the adaptation of organisms to an environment to be open-ended [Rocha, 1998]. Such adaptation to a changing environment can be seen as the selection (categorization) of organism/environment couplings into appropriate configurations for different environmental niches which under the neo-Darwinian or Von Neumann model is effectively open-ended

---

<sup>1</sup> These two distinct senses of the word “selection” should be present. The implicit selection of natural selection is quite distinct from the explicit selection of learning and cultural evolution, and lead to distinct trends of evolutionary processes, passive and active respectively [McShea, 1993, 1994]. A discussion of this distinction is provided in [Rocha, 1998c].

[Pattee, 1995; Rocha, 1996, 1997a, 1998a]. It is then at least reasonable to postulate that the apparent open-endedness of cognitive categorization may rely on the existence of some syntax or linguistic encoding which can likewise establish system a of random access to and re-creation of dynamic categories. The point here, is that language has likewise opened up a whole new universe of meaning for cognitive systems, as they can access the dynamics of classification beyond local interactions. That is, communication between individuals, as well as internally, is not restricted to only those things we can "show" or otherwise somehow physically mimic: the displacement of local observations. Language may be, as the genetic system, a method to point to and reach a particular dynamics necessary in a particular context. It may allow a (fairly) random access to an otherwise distributed memory, defining a more sophisticated system of selected self-organization [Henry and Rocha, 1996].

This view of language as a system of manipulation of an otherwise dynamic, distributed, memory expands language's role beyond a communication system among agents in selected self-organization with their environments. Its role is above all to enable the re-organization of these agent-environment dynamic couplings themselves. As discussed above, selected self-organization is based on a structural coupling between a self-organizing agent and its environment which results in the selection (categorization) of dynamic stabilities used to classify the agent-environment interaction itself. Indeed, the reality of such an agent is defined by these very stabilities or eigenbehavior [Von Foerster, 1977; Rocha, 1996] which Uexküll [1982 (1940)] named the agent's *umwelt*.

Now, this idea of language as an external mechanism of structural transformation of the dynamic machinery of the brain, rests on the establishment of some kind of linguistic code which can be shared among agents, reduces contextuality, is modality-neutral, and facilitates memorization [Clark, 1997, page 210]. Well, even though we hope to find such code in the brain, neuroscience has failed so far to identify it or even propose a viable implementation. But since a linguistic code has been unequivocally identified in the genetic machinery of biological systems, we should explore the parallels between life and cognition as much as possible to understand how can such a code ever work in the cognitive realm. It is therefore beneficial to move the categories of evolutionary discourse, particularly of theoretical biology, to cognitive science and Artificial Intelligence.

The notion of a linguistic addition to selected self-organization in cognitive systems was proposed in [Henry and Rocha, 1996; Rocha, 1997] and developed in [Rocha, 1999b], views language as an external system of structural perturbation or re-combination of distributed memory. Let us refer to this process of selected self-organization capable of re-organization through language in cognitive systems as *linguistic-based selected self-organization*. The two processes of selected self-organization, description based in biological systems, and linguistic-bases in cognitive systems lead to two distinct types of evolutionary processes, the passive evolution by natural selection and the active process of learning in cultural evolution. The distinctions between the two types of evolutionary processes are quite strong. There is indeed a need to study these distinctions [Rocha, 1998c], but it is also important to understand their fundamental similarity: the utilization of a syntactic level in the selected self-organization process of agent-environment coupling which leads to an open-ended semiosis between agent and environment. Here I am interested in exploring this open-ended semiosis as a model of cognitive categorization and as an application to information retrieval in distributed information systems.

### 1.3 Open-Ended Semiosis

Semiotics concerns the study of signs/symbols in three basic dimensions: *syntactics* (rule-based operations between signs within the sign system), *semantics* (relationship between signs and the world external to the sign system), and *pragmatics* (evaluation of the sign system regarding the goals of their users) [Morris, 1946]. Linguistic-based selected self-organization as described above manifests a full-fledged semiosis between a cognitive agent and its environment. Classification implies semantic emergence, while selection (categorization into memory) implies pragmatic environmental influence. In fact, these two dimensions of semiosis cannot be separated; the meaning of the classifications of a self-organizing system does not make sense until it is grounded in the feedback from the repercussions it triggers in its environment. The structural coupling, or situation, of a classifying, self-organizing, agent in its environment is the source of meaning. Indeed, selection does not act on memory tokens internal to a classifying system but on the repercussions those trigger in an environment. In this sense, meaning is not private to the agent but can only be understood in the context of the agent's situation in an environment with its specific selective pressures: semantics requires pragmatics.

The third dimension of semiosis, the linguistic encoding sought by Clark, must be based on some set of symbols and rules which allow significance to be transmitted and memorized categories to be recombined into new categories. In other words, it establishes a syntax. In the biological realm, a semiotic code [Umerez, 1995] is utilized to map descriptions (genetic strings) into components (aminoacid chains) which self-organize to produce some repercussion (or function) in an environment. Thus, a material semiotic code presupposes a set of components (e.g. parts and processes) for which the instructions are said to "stand for". In cognitive systems, we can only postulate that a linguistic encoding will use as components the neuronal machinery of the brain responsible for distributed memory, while the descriptions that trigger the construction and reorganization of these components are cast on whatever processes that enable and recognize linear, serially processed, linguistic syntax. The comparative leap is based on the assumption that *if* such a linguistic code exists, then cognitive categorization qua linguistic-based selected self-organization, attains the power of recombining otherwise purely dynamic categories into new ones in the same way as genetic strings (genetic categories) recombine into new proteins with description-based selected self-organization: in an open-ended manner.

Semiotics leads us to think of symbols not simply as abstract memory tokens, but as material tools [Prem, 1998] for a situated open-ended semiosis of classifying systems with their environments, which requires the definition of components that interact and self-organize with the laws of their environment [Rocha and Joslyn, 1998]. How such a semiotic code can arise from a purely dynamic self-organizing system is still very much a mystery both for biological and cognitive systems, though computational experiments to investigate the emergence of symbolic activity [Crutchfield and Mitchell, 1995; Rocha, 1998b] and even codes [Wills, 1996] have been proposed. In biology such code is clearly identified in the genetic system, but in cognition it is still very much unknown how language can access and manipulate dynamic categories. However, we should build models and systems which explore the idea of linguistic-based selected self-organization by coupling distributed memory to linguistic, random-access, manipulation of categories to try to obtain open-ended categorization of some environment. I modestly pursue this goal below.

## 2. COGNITIVE CATEGORIZATION AND EVIDENCE SETS<sup>2</sup>

Categories are bundles of classifications somehow associated in some context. Cognitive agents survive in a particular environment by categorizing their perceptions, feelings, thoughts, and language. The evolutionary value of categorization skills is related to the ability cognitive agents have to classify and group relevant events in their environments which may demand reactions necessary for their survival. If organisms can map a potentially infinite number of events in their environments to a relatively small number of categories of events demanding a particular reaction, and if this mapping allows them to respond effectively to relevant aspects of their environment, then only a finite amount of memory is necessary for an organism to respond to a potentially infinitely complex environment. The categorical flexibility of section 1.

Classifications are dynamic stabilities (attractors, eigenbehavior, and the like, as discussed above) which result from the immediate, dynamic, embodied coupling between an agent and its environment, whereas categorization is the process of grouping and memorizing such classifications. Understanding categorization as an evolutionary (control) relationship between a memory empowered organism and its environment, implies the understanding of knowledge not as a completely observer independent mapping of real world categories into an organism's memory, but rather as the organism's, embodied, thus subjective, own construction of relevant – to its survival – distinctions or classifications in its environment. Categories are pragmatically grounded and memorized classifications that can be communicated (internally or externally), and combined and re-combined linguistically to obtain new classifications of the agent-environment coupling.

### 2.1 Models of Cognitive Categories

A deeper overview of models of cognitive categories has been pursued in Rocha[1999a], for our purposes here a small overview suffices.

#### 2.1.1 The Classical View

The classical theory of categorization defines categories as containers of elements with common properties. Naturally, the classic, crisp, set structure was ideal to represent such containers: an element of a universe of observation can be either inside or outside a certain category, if it has or has not, respectively, the defining properties of the category in question. Further, all elements have equal standing in the category: there are no preferred representatives of a category – all or nothing membership.

#### 2.1.2 Prototype Theory

Rosch [1975, 1978] proposed a theory of category prototypes in which, basically, some elements are considered better representatives of a category than others. It was also shown that most categories cannot be defined by a mere listing of properties shared by all elements. Some approaches define this degree of representativeness as the distance to a salient example element of the category: a *prototype* [Medin and Schaffer, 1978]. More recently, prototypes have been accepted as abstract entities, and not necessarily a real

---

<sup>2</sup> A more complete overview of models of cognitive categorization and Evidence Sets is offered in Rocha [1999a].

element of the category [Smith and Medin, 1981]. An example would be the categorization of eggs by Lorenz' [1981] geese, who seem to use an abstract prototype element based on such attributes as color, speckled pattern, shape, and size. It is easy to fool a goose with a wooden egg if the abstract characteristics of the prototype are emphasized.

### 2.1.3 Dynamic Categories

As Hampton [1992] and Clark [1993] discuss, the important question to ask at this point is “where do the prototypicality degrees come from?” Barsalou [1987] has shown how the prototypical judgments of categories are very unstable across contexts. He proposes that these judgements, and therefore the structure of categories, are constructed “on the hoof” from contextual subsets of information stored in distributed long-term memory. The conclusion is that the wide variety of context-adapting categories we use cannot be stored in our brains, they are instead dynamic categories which are rarely, if ever, constructed twice by the same cognitive system. Categories have indeed Rosch's graded prototypicality structure, but they are not stored as such, merely constructed “on the hoof” from some other form of information storage system.

As Clark [1993] points out, the reason for this is that since the evidence for graded categories is so strong, even in ad hoc categories such as “things that could fall on your head” or viewpoint-related categories, “it seems implausible to suppose that the gradations are built into some preexisting conceptual unit or prototype that has been simply extracted whole out of long-term memory.” [Ibid, page 93] Thus, we should take the graded prototypical categories as representations of these highly transient, context-dependent knowledge arrangements, and not of models of information storage in the brain.

## 2.2 Sets and Cognitive Categorization: the *Prototype Combination Problem*

Fuzzy sets<sup>3</sup> [Zadeh, 1965] are fairly accurate representations of categories because they are able to represent prototypicality (understood as degree of representativeness); how the prototype degrees are constructed is, on the other hand, a different matter. Fuzzy sets are simple representations of categories which need much more complicated models of approximate reasoning than those fuzzy set theory alone can provide in order to satisfactorily model cognitive categorization processes. Critics [Osherson and Smith, 1981; Smith and Osherson, 1984; Lakoff, 1987] have shown that the several fuzzy set connectives (e.g. intersection and union) cannot conveniently account for the prototypicality of the elements of a complex category. This is known as the *prototype combination problem*.

A complex category is assumed to be formed by the connection of several other categories. Smith and Osherson's [1984] results, showed that a *single* fuzzy connective cannot model the association of entire categories into more complex ones. Their analysis centered on the traditional fuzzy set connectives of (max-min) union and intersection. They observed that max-min rules cannot account for the membership degrees of elements of a complex category which may be lower than the minimum or higher than the maximum of their membership degrees in the constituent categories. Their analysis is very incomplete regarding the full-scope of fuzzy set connectives, since we can use other operators [see Dubois and Prade, 1985], to obtain any desired value of membership in the  $[0, 1]$  interval of membership. However, their basic criticism remains valid: even if we find an appropriate fuzzy set connective for a particular element, this connective will not

---

<sup>3</sup> Elements are included in the set with a membership degree between 0 (not a member) and 1 (full membership).

yield an accurate value of membership for other elements of the same category. Hence, a model of cognitive categorization which uses fuzzy sets as categories will need several fuzzy set connectives to associate two categories into a more complex one (in the limit, one for each element). Such model will have to define the mechanisms which choose an appropriate connective for *each* element of a category. No single fuzzy set connective can account for the exceptions of different contexts, thus the necessity of a complex model which recognizes these several contexts before applying a particular connective to a particular element.

## 2.3 Beyond Direct Association and Distributed Memory

The prototype combination problem is not only a problem for fuzzy set models, but for all models of combination of prototype-based categories. Fodor [1981] insists that though it is true that prototype effects obviously occur in human cognitive processes, such structures cannot be fundamental for complex cognitive processes (high level associations): “there may, for example, be prototypical cities (London, Athens, Rome, New York); there may even be prototypical *American Cities* (New York, Chicago, Los Angeles); but there are surely not prototypical *American cities situated on the east coast just a little south of Tennessee.*” [Ibid, page 297] As Clark [1993] points out, the problem with Fodor’s point of view, and indeed also the reason why fuzzy set combination of categories fails, is that “he assumes that prototype combination, if it is to occur, must consist in the linear addition of the properties of each contributing prototype.” [Ibid, page 107] Clark proposes the use of connectionist prototype extraction as an easy way out of this problem. In fact, a neural network trained to recognize certain prototype patterns, e.g. some representation of “tea” and “soft drink”, which is also able to represent a more complex category such as “ice tea”, “does not do so by simply combining properties of the two ‘constituent’ prototypes. Instead, the webs of knowledge structure associated with ‘hot spots’ engage in a delicate process of mutual activation and inhibition.” [Ibid, page 107] In other words, complex categories are formed by nonlinear, emergent, prototype combination.

As Clark points out, however, this ability to nonlinearly combine prototypes in connectionist machines is a result of the pre-existence of a (loosely speaking) *semantic metric* which relates all knowledge stored in the connectionist network. Through the workings of the network with its inhibition and activation signals, new concepts can be learned which must somehow relate to the existing knowledge previously stored. Therefore, any new knowledge that a connectionist device gains, must be somehow related to previous knowledge. This dependence on direct association prevents the sort of open-ended semiosis, or concept recombination, that we require of linguistic-based selected self-organization.

This problem might be rephrased by saying that connectionist devices can only make nonlinear prototype combinations given a small number of contexts. The brain may use a network to classify, say, sounds, another one images, and so forth. In their own contexts, each network combines prototypes into more complex ones, but they cannot escape their own contexts. A computational model is presented in section 3 to deal with this contextual problem. This model even though not using connectionist machines or distributed memory in the strong sense of nonlinear superposition of categories [van Gelder, 1992], uses networked databases possessing structural and semantic semi-metrics that constantly engage in (Hebbian type) cross-activation and inhibition of links between tokens of knowledge. Networked information with changing structural and semantic semi-metrics is distributed in the weaker sense of mere network association of localized (not superposed) knowledge tokens. However, to move beyond purely associative distributed semantic semi-metrics, this model relies on a higher-level linguistic coupling to an environment designed to achieve the categorical recombination desired for open-ended semiosis as discussed in section 1. Details in section 3, but first we need to define the components of the linguistic coupling: evidence sets.



## 2.4 Evidence Sets: Contextual Categories

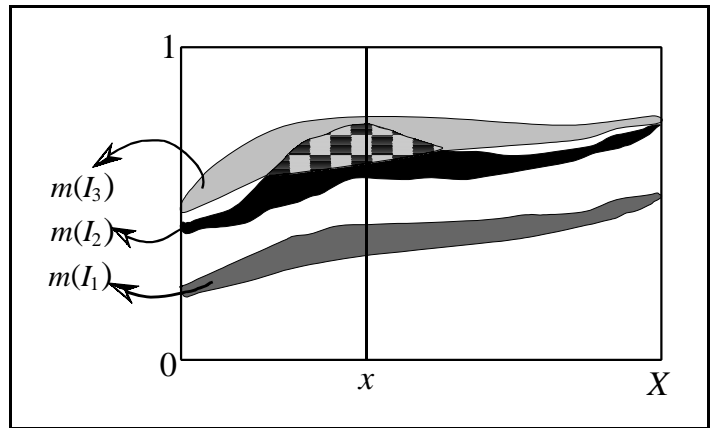
The problem with fuzzy sets as models of cognitive categories lies on them lacking an explicit mechanism to account for context dependencies and to deal with the subjective judgments of categorizing cognitive systems. This issue is discussed in detail in [Rocha, 1997a, 1999a]. To overcome these limitations a new set structure referred to as *evidence set* was introduced [Rocha, 1994, 1997a, 1999a], which extends fuzzy sets with the Dempster-Shafer Theory of Evidence (DST) [Shafer, 1976]. To explain what an evidence set is, let us start with fuzzy sets (for a more mathematical description please refer to Rocha [1997a, 1997b, 1999a]).

The membership degree of an element in a fuzzy set is given by a real value in the unit interval: 0 denotes non-membership and 1 maximum membership. This degree introduces uncertainty in the definition of set membership: instead of an element simply being or not being a member of a set, as crisp sets demand, an element is a member of a set to a degree, which conversely implies that the element is *also* not a member of the set to the reciprocal degree. This kind of uncertainty – simultaneously being and not being to a degree – is referred to as *fuzziness*.

A different kind of uncertainty is introduced when we allow the degree of membership to be represented by a subinterval of the unit interval. Now instead of a specific degree we have a nonspecific interval to represent a degree of membership of an element in a set. This kind set is called an interval valued fuzzy set (IVFS) whose degrees of membership capture two types of uncertainty: fuzziness and *nonspecificity*.

The membership degree of an element in an evidence set is defined by a probability restriction on a collection of subintervals of the unit interval. This probability restriction on subsets (not elements) is defined by the DST (details in Rocha [1997a, 1997b, 1999a]). Each interval of membership is associated with a weight, and the collection of weights must add to one (a probability restriction). Now the membership representation is divided into several, possibly disjoint, subintervals which means that we have conflicting evidence as to where the actual degree of membership lies. This way the membership representation of evidence sets introduces a third kind of uncertainty: *conflict*. Details about the uncertainty content of evidence sets and how to measure it are provided in Rocha[1997b].

The interpretation I suggest for the multiple intervals of evidence sets, defines each interval of membership with its correspondent evidential weight as the representation of the prototypicality of a particular element in category *according* to a particular *perspective*. Thus, the membership of each element of an evidence is defined by several intervals representing different, possibly conflicting, perspectives. The ability to maintain several of these perspectives, which may conflict at times allows a model of cognitive categorization or knowledge representation to *directly* access particular contexts influencing the definition of a particular category. In other words, the



**Figure 1:** Non-consonant evidence set with 3 focal elements:  $m(I_1) + m(I_2) + m(I_3) = 1$ .

several intervals of membership of evidence sets refer to different perspectives which *explicitly* point to particular contexts. Figure 1 depicts an evidence set with 3 disjoint focal intervals.

Evidence sets are set structures which provide interval degrees of membership, weighted by the probability constraint of DST. IVFS, fuzzy sets, and crisp sets are all special cases of evidence sets: an IVFS denotes a set with nonspecific and fuzzy membership with one single perspective, a fuzzy set denotes a set with fuzzy membership and one single perspective, and a crisp set denotes a set with a single perspective and no uncertainty in membership. The basic set operations of complementation, intersection, and union have been defined and establish a belief-constrained approximate reasoning theory of which fuzzy approximate reasoning and traditional set operations are special cases [Rocha 1997a, 1997b, 1999a].

Evidence Sets are defined by two complementary dimensions: membership and belief. The first represents a fuzzy, nonspecific (interval-based), degree of membership, and the second a subjective degree of belief on that membership. The subjective nature of DST was advanced by Shafer [1978], who proposed the values of the theory's probability constraint as *judgements*, formalized in the form of a degree [Shafer, 1976, page 21]. For more details on the nature of this subjective interpretation and DST, please refer to Rocha[1997a, 1999a]. Likewise, Rosch's prototypicality is not meant to be an objective grading of concepts in a category, but rather judgements of some uncertain, highly context-dependent, grading [Rosch, 1978, page 40]. Evidence sets offer a way to model these ideas since a membership grading (with a full account of uncertainty effects) of elements in a category is offered together with an explicit formalization of the belief posited on this membership. For evidence sets, membership in a category and judgments qualifying this membership are different, complementary, qualities of prototypicality. For a deeper discussion of the mathematics of Evidence Sets and their ability to model cognitive categories please refer to [Rocha et al, 1996; Rocha, 1997a, 1997b, 1999a].

Though based on two dimensions of membership and belief, the combination of two evidence sets into a new one as defined by the operations of belief-constrained approximate reasoning relies solely on the information present on the progenitor pair and the linear combination of this information. To overcome the prototype combination problem and to model more appropriately the generation of cognitive categories as previously discussed, a hybrid system of distributed information and belief-constrained approximate reasoning is proposed next. This system models aspects of cognitive categorization as a linguistic-based selected self-organization process, and is above all a useful tool for adaptive recommendation which enables an open-ended semiosis between distributed information systems and their communities of users.

### **3. ADAPTIVE RECOMMENDATION FOR DISTRIBUTED INFORMATION SYSTEMS**

#### **3.1 Distributed Information Systems and Information Retrieval<sup>4</sup>**

Distributed Information Systems (DIS) refer to collections of electronic networked information resources in some kind of interaction with communities of users; examples of such systems are: the Internet, the World Wide Web, corporate intranets, databases, library information retrieval systems, etc. DIS serve large and diverse communities of users by providing access to a large set of heterogeneous electronic information

---

<sup>4</sup> This subsection stems from an essentially "nonlinear" collaboration with Cliff Joslyn at the Los Alamos National Laboratory. Many of the ideas described are undoubtedly due to him.

resources. As the complexity and size of both user communities and information resources grows, the fundamental limitations of traditional information retrieval systems have become evident.

Information Retrieval refers to all the methods and processes for searching relevant information out of information systems (e.g. databases) that contain extremely large numbers of documents. Traditional information retrieval systems are based solely on *keywords* that index (semantically characterize) documents and a query language to retrieve documents from centralized databases in terms of these keywords. This setup leads to a number of flaws:

- ***Passive Environments.*** There is no genuine interaction between user and system, the former *pulls* information from a passive database and therefore needs to know how to query relevant information with appropriate keywords. Furthermore, such impersonal interfaces cannot respond to queries in a user-specific fashion because they do not keep user-specific information, or user profiles. The net result is that users must know in advance how to characterize the information they need before pulling it from the environment.
- ***Idle Structure.*** Structural relationships between documents, keywords, and information retrieval patterns are not utilized. Different kinds of structural relationships are available, but not typically used, for different DIS, e.g. citation structure in scientific library databases, the link structure in the WWW, the clustering of keyword relationships into different meanings of keywords, temporal patterns of retrieval, etc.
- ***Fixed Semantics.*** Keywords are initially provided by document authors (or publishers, librarians, and indexers), and do not necessarily reflect the evolving semantic expectations of users.
- ***Isolated Information Resources.*** No relationships are created or information is exchanged among documents and/or keywords in different information resources such as databases, web sites, etc. Each resource is accessed with a private set of keywords and query language.

These flaws prevent current information retrieval processes in DIS to achieve any kind of interesting coupling with users. No system-user semiosis can be achieved because of the following fundamental limitations:

- There is no ***recommendation***. Because of passive environments and idle structure, information retrieval systems cannot pro-actively push relevant information to its users about related topics that they may be unaware of.
- There is no ***conversation*** between users and information resources, between information resources, and between users. Because of passive environments and isolated information resources there is no mechanism to exchange knowledge, or crossover of relevant information.
- There is no ***creativity***. Because of fixed semantics, isolated information resources, idle structure, and passive environments, there is no mechanism to recombine knowledge in different information resources to infer new categories of keywords used by different communities of users.

### 3.2 Active Recommendation Systems

New approaches to information retrieval have been proposed to address these limitations. *Active recommendation systems*, also known as *Active Collaborative Filtering* [Chislenko, 1998], *Knowledge Mining*, or *Knowledge Self-Organization* [Johnson et al, 1998] are information retrieval systems which rely on active computational environments that interact with and adapt to their users. They effectively push relevant information to users according to previous patterns of information retrieval or individual user profiling.

Recommendation systems are typically based on user-environment interaction mediated by intelligent agents or other decentralized components and come in two varieties [Balabanović and Shoham, 1997]:

- In *content-based* recommendation, user profiles are created based on the system's keywords. These establish a means of recommending documents to users according to their profiles and some kind of semantic metric that describes the relationships between keywords inferred from their association with common documents.
- In *collaborative* recommendation no description of the semantics or content of documents is involved, rather recommendations are issued according to a comparison of the profiles of several users that tend to access the same documents. These user profiles are not based on keywords, but on the actual documents retrieved.

Content-based systems depend on single user profiles, and thus cannot effectively recommend documents about previously unrequested content. Conversely, pure collaborative systems, with no content analysis, match only the profiles of users that (to a great extent) have requested the same exact documents; for instance, different book editions or movie review web sites from different news organizations are considered distinct documents. It is clear that effective recommendation systems require aspects of both approaches.

Hybrid approaches to recommendation usually rely on software agents and a central database. The agents have two distinct roles:

1. to retrieve and collect documents from information resources into a database or router
2. to select or filter those documents retrieved which match the profile of specific users.

This is the case, for instance, of *Fab* [Balabanović and Shoham, 1997] and *Amalthaea* [Moukas and Maes, 1998]. These systems clearly establish active environments which are capable of recommendation, that is, they push topics that users may have not thought of, rely on user-specific interfaces that enable personalized user-environment interaction, and keep track of historical data of this interaction. In the terms used above, these systems expand information retrieval beyond passive environments and completely idle structure (they keep track of user-environment interaction).

From the picture of information retrieval depicted in 3.1, there is clearly still much more room to improve. The structure of DIS is still largely idle in these collaborative systems. Indeed these systems can improve considerably by clustering and ranking documents according to the structure of keyword relationships [Kannan and Vempala, 1999] or the structure of document linkage [Kleinberg, 1998]. These data-mining and graph-theoretical improvements can and should be used to move beyond idle structure of information retrieval in DIS and achieve a much more powerful recommendation capability.

In the present work, however, the goal is to improve recommendation systems by empowering them with effective conversation and creativity dimensions as described in 3.1. For this we need to develop more active environments and move beyond fixed semantics and isolated information resources, while keeping up the tapping of the no longer idle structure of DIS. In the following the *TalkMine* system is described which provides a mechanism to exchange and recombine knowledge between users and information resources, as well as among information resources themselves.

### 3.3 Adaptive Recommendation with *TalkMine*: Computing Categories through Conversation in DIS

*TalkMine* is an adaptive recommendation system which is both collaborative and content-based, and exploits currently untapped sources of information in DIS. In particular, it integrates information from the patterns of usage of groups of users, and also categorizes DIS content or semantics in a manner relevant to those groups. Moreover, the keywords and derived categories need not be just designed into these systems, but are also induced and evolved from document content, user-supplied information, and group interaction. To show how conversation and creativity are enabled and establish an open-ended semiosis between user and DIS mediated by *TalkMine*, this system is described in detail below. A discussion of the system regarding the notion of linguistic-based selected self-organization is provided in section 4 where the connection to Pask's work is also made explicit.

*TalkMine* is a conversational recommendation system for DIS that uses evidence sets as categorization mechanisms. It defines a human-machine interface that can capture more efficiently the user's interests through an interactive question-answering process. It also models certain aspects of cognitive categorization processes as linguistic-based selected self-organization. The model offers an expansion of Nakamura and Iwai's [1982] data-retrieval system which is expanded from a fuzzy set to an evidence set framework. The evidence set expansion allows the construction of categories from several information resources simultaneously.

#### 3.3.1 The Distributed Memory Structure

Each information resource (e.g. a database) is defined by a network structure with two different kinds of objects: *Keywords*  $x_i$  and *Documents*  $n_i$  (e.g. data records like books, web pages, etc). Each keyword semantically classifies or indexes a number of documents which may be shared with other keywords. These two different kinds of objects establish a *structure* and a *semantics* of the information resource.

The structure of an information resource refers to the information that can be obtained by the relationships between documents alone, which are formalized by a graph whose nodes are documents and the edges the relationships between documents, e.g. hyperlinks between web pages or citation structure in databases of academic documents. The edges of this graph in many cases are formalized by a weight  $w_{ij} \in [0, 1]$  which denotes some relevant normalized strength of relationship between two documents  $n_i$  and  $n_j$  (e.g. ratio of links out of a web page). To discern the closeness of documents, we can now define measures of proximity between any two documents.

Let  $p_{i,j}^{out}$  denote the *outwards direct proximity* between any two document nodes  $n_i$  and  $n_j$ :

$$p_{i,j}^{out} = \frac{\sum_{k \in N_d^{i,j}} \min(w_{i,k}, w_{j,k})}{\sum_{k \in N_d^{i,j}} \max(w_{i,k}, w_{j,k})} \quad (1)$$

where  $N_d^{i,j}$  represents the set of nodes linked from nodes  $n_i$  and  $n_j$  up to a depth of level  $d$ . Let  $p_{i,j}^{in}$  denote the *inwards direct proximity* between any two nodes  $n_i$  and  $n_j$ :

$$p_{i,j}^{in} = \frac{\sum_{k \in N_d^{i,j}} \min(w_{k,i}, w_{k,j})}{\sum_{k \in N_d^{i,j}} \max(w_{k,i}, w_{k,j})} \quad (2)$$

These proximity measures<sup>5</sup> may be calculated globally for the entire information resource, or locally by each node of the network up to some desired depth, if the size of the information resource demands a distributed approach. The outwards (inwards) proximity relates any two nodes according to the number of nodes both have outwards (inwards) edge links to (from). Their values vary in the unit interval. A non-directional value of proximity between any two nodes can be obtained by linear combination (e.g. averaging) of the values of (1) and (2) for these nodes. From this value we can define a *neighborhood* of a node  $n_i$  as the set of nodes related to node  $n_i$  with proximity greater than  $\alpha \in [0, 1]$ .

Many structural properties of information resources may be obtained from the graph of documents and the proximity measures, such as clustering of nodes and the study of relative importance of documents [Kleinberg, 1998]. For our purposes here we require only the ability to compute the neighborhood of documents or the proximity information in some subset of the overall graph structure.

The semantics of an information resource is also formalized by a graph whose nodes are keywords, and denotes the relationships between keywords alone, which are (at first) obtained from document-keyword relationships. Based on the amount of documents shared with one another, a measure of *semantic proximity*,  $s$ , can be constructed for keywords  $x_i$  and  $x_j$ :

$$s(x_i, x_j) = \frac{N(x_i \cap x_j)}{N(x_i \cup x_j)} = \frac{N(x_i \cap x_j)}{N(x_i) + N(x_j) - N(x_i \cap x_j)} \quad (3)$$

where  $N(x_i)$  represents the number of documents that are directly indexed by keyword  $x_i$ ,  $N(x_j)$  represents the number of documents that are directly indexed by keyword  $x_j$ , and  $N(x_i \cap x_j)$  represents the number of

---

<sup>5</sup>Equations (1) and (2) are proximity measures, as they establish reflexive ( $p_{i,i} = 1$ ) and symmetrical ( $p_{i,j} = p_{j,i}$ ) values among the nodes of the network. If the law of transitivity is additionally observed, then (1) and (2) are also *similarity* or *equivalence* relations.

documents that are directly indexed either  $x_i$  or  $x_j$ . The inverse of the semantic proximity,  $s$ , defines a measure of distance<sup>6</sup>,  $d$ :

$$d(x_i, x_j) = \frac{1}{s(x_i, x_j)} \quad (4)$$

(4) may be applied to the keyword-document structure of the whole information resource, or only to a given structural neighborhood induced from the structural proximity of (1) and (2), or to some other subset of the overall structure. The distances between directly linked keywords are calculated using (4). After this, the shortest path is calculated between indirectly linked keywords. The algorithm allows the search of indirect distances up to a certain level. The set of *n-reachable* keywords from keyword  $x_i$ , is the set of keywords that have no more than  $n$  direct paths between them. If we set the algorithm to paths up to level  $n$ , all keywords that are only reachable in more than  $n$  direct paths from  $x_i$  will have their distance to  $x_i$  set to infinity.

A *Local Knowledge Context*  $X_k$  denotes the semantics of information resource  $k$  or one of its subsets such as the neighborhood of a given document. It is the set of all its keywords and the matrix of their relative distance  $d_k$  as constructed with the semi-metric (4). It captures the knowledge of an information resource by keeping a record of associations between keywords, as well as a measure of their semantic similarity. Notice that the semantic proximity information is abstracted from the document-keyword relationship and is not stored as such in the document structure. There is a parallel here to connectionist devices. Clark [1993] proposed that connectionist associative devices work by producing some kind of semantic metric which relates the knowledge they store. This metric and the knowledge it relates are not stored locally in the nodes of the network, but rather nonlinearly superposed over its weights [van Gelder, 1992].

A local knowledge context is not a connectionist structure in a strong sense since the keywords can be identified in particular nodes in the network. However, the same keyword is found in many nodes of the document structure. Losing a few document nodes will not affect significantly the derived semantic metric for a large enough network. In this sense, keywords are distributed over the entire network of document nodes in a highly redundant manner as required of sparse distributed memory models [Kanerva, 1988], and observe the semantic categorical flexibility discussed in 1. Furthermore, the semantic proximity is abstracted from this distributed document-keyword structure, thus constructing global associative information which is not stored as such in any one location of the network. The global associative information (the semantic metric) is constructed from the integration of the contribution of many components which is akin to the process of self-organization described in 1. Therefore, a local knowledge context does possess a global semi-metric space generated from keyword information distributed in a network of nodes. This semantic semi-metric associates the knowledge stored, as desired of connectionist engines [Clark, 1993]. Below I discuss how this semantic associative metric adapts to users (the environment) with Hebbian type learning.

A local knowledge context is the lower-level structure of an information resource; it instantiates its long-term memory banks. Its semantic metric is unique, and reflects the semantic relationships established by the set

---

<sup>6</sup> This measure of distance calculated in a large network of nodes, is usually not a Euclidean metric because it does not observe the triangular inequality. In other words, the shortest distance between two nodes of the network might not be the direct path. This means that two nodes may be closer to each other when another node is associated with them. Such measures of distance are referred to as semi-metrics [Galvin and Shore, 1991].

of documents stored. Thus, the semantic semi-metric defined by (4), reflects the actual inter-significance of keywords for the system and its community of users and authors. The same keyword in different information resources will be related differently to other keywords, because the actual documents stored will be distinct. The documents stored in an information resource are a result of the history of utilization and deployment of information by its authors and users. Thus, each local knowledge context captures the knowledge that the community of users of the deriving information resource has accumulated in some context.

The *Total Knowledge Space*  $X$  of a collection of  $n_d$  information resources is the set of all local knowledge contexts associated with these information resources:

$$X = \bigcup_{k=1}^{n_d} X_k$$

This system has  $n_d$  different distance semi-metrics,  $d_k$ , associated with each information resource. Since each of the  $n_d$  information resources has a different keyword-document pattern of connectivity, each distance semi-metric  $d_k$  will be different. If the local knowledge spaces derive from similar contexts, naturally their distance semi-metrics will tend to be more similar. From this collection of semantic semi-metrics associated with different contexts, that is, from the long-term distributed memory of a DIS, temporary prototype categorizations can be formed to model the “on the hoof” categories previously discussed.

### 3.3.2 Short Term Categorization Processes

With their several intervals of membership weighted by a probability restriction, Evidence sets can be used to quantify the relative interest of users in each of the local knowledge contexts from the  $n_d$  information resources. This relative interest may also be automatically generated by clustering and ranking of keywords the user is interested in the several information resources [Kannan and Vempala, 1999]. *TalkMine* uses a question-answering process to capture the user’s interests in terms of the system’s own long-term distributed memory. In other words, the system constructs its own internal categories in interaction with the community of users. This user-system conversation is implemented using the evidence set operations of intersection and union [see Rocha, 1997a, 1999a for details].

The system starts by presenting the several information resources available to the user, who has to probabilistically grade them (weights must add to one) or choose an automatic grading given a number of keywords or interests previously selected. The selected information resources define the several local knowledge contexts which the system uses to construct its categories. The question-answering algorithm is defined as follows:

1. The user selects the  $n_d$  information resources of interest and their respective weights  $m_k$ .
2. The user inputs an *initial keyword of interest*  $x_i \in X$ .
3. The system creates an evidence set membership function centered on  $x_i$  and affecting all its close neighbors proportionally to their semantic proximity in each of the local knowledge contexts. This resulting evidence set of  $X$  represents a category that keeps the user’s interests in terms of the system’s own associative information: *The learned category*  $A(x)$ .
4. The system calculates the total uncertainty of the learned category in its forms of fuzziness, nonspecificity, and conflict (the uncertainty measures used are discussed in



- Rocha[1997b]). If total uncertainty is below a pre-defined small value the process stops, otherwise continue.
5. The system selects another keyword  $x_j \in X$ .  $x_j$  is selected in order to potentially minimize the uncertainty of the learned category.
  6. The user is asked whether or not she is interested in  $x_j$ .
  7. If the answer is “YES” another membership function as created in step 3 is created over  $x_j$ , and an evidence set union is performed with the previous state of the learned category.
  8. If the answer is “NO” the inverse of the membership created in 7 is created over  $x_j$ , and an evidence set intersection is performed.
  9. The system calculates the total uncertainty of the learned category in its forms of fuzziness, nonspecificity, and conflict.
  10. If the uncertainty of the learned category is smaller than half the maximum value attained previously, the system stops since the learned category is considered to have been successfully constructed. Otherwise computation goes back to step 5.

Several approaches can be used to define evidence set membership functions for the algorithm above. The precise functions used by *TalkMine* are described in Rocha[1997a, 1999a]. It is important to notice that the evidence set categories constructed with this algorithm, are not stored in any location in the long-term distributed memory. They are temporarily constructed by integration of long-term knowledge from several information resources and the interests of the user expressed in the interactive conversational process. These constructed categories are therefore temporary containers of knowledge nonlinearly integrated from and relevant for the user and the collection of information resources. Thus, this algorithm implements many of the, temporary, “on the hoof” [Clark, 1993] category constructions ideas as discussed previously. In particular, it is based on a long-term distributed memory bank of associations that implements the system’s own semantic relationships. Prototype categories are then built using evidence sets which reflect both such contextually dependent semantic metrics and the directed interest of a user.

Notice that since each information resource spawns different local knowledge contexts with associated different semi-metrics, the learned categories constructed with the above algorithm may possess the three types of uncertainty discussed in section 2, because keywords will be associated in different ways for each local knowledge context. For instance, two keywords highly associated in one context, may be highly unrelated in another one which implies the existence of a conflict between the semantics of each context. It is important to stress that this more accurate construction of prototypical categories includes uncertainty forms as a result of semantic differences in the information stored in the several long-term distributed memory banks utilized. It is the lower level uncertainty/conflict of the long-term memory resources that is reflected in the short-term construction of categories by conversation with users reflects.

### 3.3.3 Document Retrieval

After construction of the final learned category, the system must return to the user documents relevant to this category. Notice that every document  $n_i$  defines a crisp subset of the knowledge space  $X$  whose elements are all the keywords which characterize  $n_i$  in all the constituent information resources. Since each document defines a crisp subset of  $X$ , the similarity between this crisp subset and the evidence subset defined by the learned category is a measure of the relevance of the property to the learned category. This similarity may be defined by different ways of calculating the subethood of one subset in the other. Details of the actual

operations used are presented in Rocha[1999a]. High values of these similarity measures will result on the system returning only those documents highly related to the learned category.

### 3.3.4 Adaptive Alteration of Long-Term Memory by Short-Term Categorization

The final component of *TalkMine*, which makes it an instance of linguistic-based selected self-organization, is the adaptation of the long-term distributed memory to the community of users of this system – effectively the system’s environment. Due to the structure of information resources (the documents stored), the derived semantic semi-metrics may fail to construct associations between keywords that their users find relevant. Furthermore, most documents in a given information resource do not change (e.g. scientific articles), producing a fixed semantics as discussed in section 3.1. But the semantics of users change with time as new keywords and associations between keywords are created and changed. Therefore, an effective recommendation system for DIS needs to adapt its semantic associations to the evolving semantics of its users.

The scheme used to implement this adaptation is very simple: the more certain keywords are combined with each other, by often being simultaneously included with a high degree of membership in the learned categories that result from the algorithm in section 3.3.2, the more the semantic distance between them is reduced. Conversely, if certain keywords are not frequently associated with one another, the distance between them is increased. An easy way to achieve this is to have the values of  $N(x_i)$  and  $N(x_i \cap x_j)$  as defined in (3), adaptively altered for each of the constituent  $n_d$  information resources. After an evidence set learned category is constructed and approximated by a fuzzy set  $A(x)$ , these values are changed according to:

$$N_{t+1}^k(x_i) = N_t^k(x_i) + A(x_i), \quad k=1\dots n_d, \quad x_i \in X \quad (5)$$

and

$$N_{t+1}^k(x_i \cap x_j) = N_t^k(x_i \cap x_j) + \min[A(x_i), A(x_j)], \quad k=1\dots n_d, \quad x_i \in X \quad (6)$$

respectively ( $t$  indicates the current state and  $t+1$  the new state). This implements an adaption of the long-term distributed memory of information resources to their users according to repeated association of keywords in categories constructed in conversation with users. This adaptation leads the associative semantic semi-metric of the local knowledge contexts involved to increasingly match the expectations of the community of users with whom the system interacts. In other words, the long-term distributed memory is consensually selected by the community of users: we observe the selected self-organization of knowledge stored in the information resources coupled to an environment of users and other information resources.

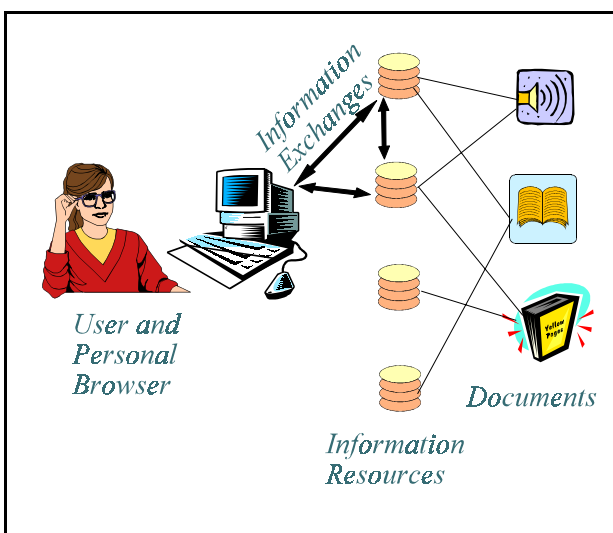
Furthermore, when highly activated keywords in the learned category are not present in the same information resource (existing in some other information resources combined simultaneously in a given user conversation) they are added to the information resource which does not contain them with property counts given by equations (5) and (6). If the simultaneous association of the same keywords keeps occurring, then an information resource that did not previously contain a certain keyword, will have its presence progressively strengthened, even though such keyword does not really characterize any documents stored in this information resource.

### 3.3.5 TalkMine: The Implemented Application

*TalkMine* was initially developed as a prototype application for personal computers, see Rocha[1997a, 1999a] for details. Currently, it is being developed as a testbed environment for the Research Library at the *Los Alamos National Laboratory*, more specifically, for its *Library Without Walls project*<sup>7</sup>. With this implementation *TalkMine* will use as information resources several of the extensive electronic databases available to this research library.

The architecture of *TalkMine* has both user-side and system-side components. Each user owns a browser (or plug-in to an existing Internet browser), which functions as a consolidated interface to all information resources searched. This individual browser stores user preferences and tracks information retrieval patterns and relationships which it utilizes to adapt to the user. User preferences are stored as a set of local knowledge contexts which the user has constructed while using the system under a set of different interests. These local knowledge contexts store both semantic semi-metric and structural proximity information. This way, user preferences are much more than a list of keywords used or documents retrieved (e.g. a list of “Bookmarks”), because they also keep ever adaptive associative information between keywords and between documents. In other words, the browser keeps track of the semantic inter-associations that have been relevant for the user. This training can be done for distinct sets of user interests, that is, the user can choose to train its browser when she retrieves information as, say, a scientist or as a sports aficionado. Each of the associated local knowledge contexts can be seen as a sort of surrogate “personality” which can be used to automate the question-answering process of section 3.3.2. In others words, the browser can participate in this conversation in lieu of the user who trained it. All user profiles are stored in the user’s browser and do not need to be transmitted to the information resources, except as yes or no answers to the questions in the conversation process of section 3.3.2. Indeed, the adaptation of the information resources does not require any personal information to instantiate the rules of section 3.3.4, which makes this architecture of user-side and system-side components secure and private.

Where existing information retrieval is strictly unidirectionally query-based, in *TalkMine* an interactive, *conversational*, multi-directional approach between user and system side components is fundamental. Each user’s browser engages in the interactive algorithm of section 3.3.2 with the information resources it queries. This first results in a list of document and related topic recommendations issued according to the user’s profile and present interests, as well as the integration of knowledge from the several information resources queried, as discussed above. The second result of this interaction is that all sides exchange information, therefore all of the parties can potentially learn new information in an adaptive fashion. Indeed, information resources can learn new keywords from users and other information



**Figure 2:** TalkMine establishes an adaptive system of information among the user’s browser and the several information resources it queries.

<sup>7</sup> More details of this project at <http://www.c3.lanl.gov/~rocha/lww>.

resources, and will adapt the associations between keywords and documents according to the expectations of its users.

*TalkMine* is an adaptive recommendation system which is both content-based and collaborative. It tackles the flaws of information retrieval in DIS as depicted in section 3.1 in the following manner:

- It establishes an *active environment* of user-system interaction capable of recommending information relevant to the particular users and the expectations of the overall community of users.
- It explores *structural relationships* in the document structure with the proximity measures of section 3.3.1. Further exploitation of structural relationships can be achieved with many data-mining techniques [e.g. Kannan and Vempala, 1999], which future developments of *TalkMine* will employ, but this system goes well beyond the idle structure of traditional information retrieval in DIS in its current layout.
- It establishes an *evolving semantics* as keyword associations adapt to the expectations of users and new keywords are introduced from the crossover of information among multiple information resources and users browsers.
- It establishes *linked information resources* as users can use personal browsers to search several resources simultaneously and establish all-way information exchanges.

Therefore, *TalkMine* overcomes the limitations of information retrieval outlined in 3.1:

- There is *recommendation* as the system pro-actively pushes relevant documents to users about related topics that they may have been unaware of. This is achieved because of the structural and semantic proximity information kept in the distributed memory (section 3.3.1), its integration with user-specific (also structural and semantic) information in the categorization process (section 3.3.2), and finally by the document retrieval operations (section 3.3.3).
- There is *conversation* between users and information resources and among information resources (and indirectly among users) as a mechanism to exchange or crossover knowledge among them is established. As categories are constructed with the question-answering process (section 3.3.2), a list of documents is produced (section 3.3.3) and communicated not only to users but also to information resources that did not contain them, and the semantics of all parties involved are adapted (section 3.3.4).
- There is *creativity* as new semantic and structural associations are set up by *TalkMine*. The categorization process brings together knowledge from the different contexts of the information resources. This not only adapts existing local semantics, but combines knowledge not locally available to individual information resources. In this sense, because of the conversation process, information resources gain *new* knowledge previously unavailable.

For all of these characteristics, *TalkMine* establishes an open-ended human-machine symbiosis, which can be used in the automatic, adaptive, organization of knowledge in DIS such as library databases or the Internet, facilitating the rapid dissemination of relevant information and the discovery of new knowledge. More on this open-ended semiosis in section 4.

## 4 PASK AND OPEN-ENDED SEMIOSIS

### 4.1 Pask: Selprunes as “on the hoof” Categories

The initial development of *TalkMine* started from an attempt to re-create aspects of Gordon Pask’s [1975] Conversation Theory with different formal tools such as fuzzy logic and distributed memory, and a re-phrasing of its constructivist position in current-day cognitive science discourse [Rocha, 1997a]. The hope is that this re-formulation and re-interpretation can both develop and secure the longevity of his views. Part of the work discussed in this section was first presented to Gordon Pask in 1991 under the heading of “Fuzzification of Conversation Theory”<sup>8</sup>. Evidence Sets, the extensions

of fuzzy sets described earlier, were developed precisely from some of Gordon’s criticisms of the work then presented, namely that with fuzzy sets alone there is no explicit context representation, and by his posing of the “where do fuzzy degrees come from?” question discussed earlier (section 2.1.3). These criticisms lead me precisely to the development of the hybrid architecture of *TalkMine* that uses connectionist-like long term distributed memory and short-term categorizations as a conversational, linguistic, mechanism to allow an effective coupling of long-term memory to the environment (users and other information resources) which can re-organize the memory in a more effective manner than pure connectionist devices.

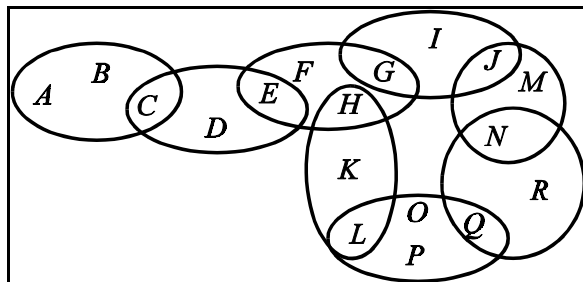


Figure 3: An example of an entailment mesh

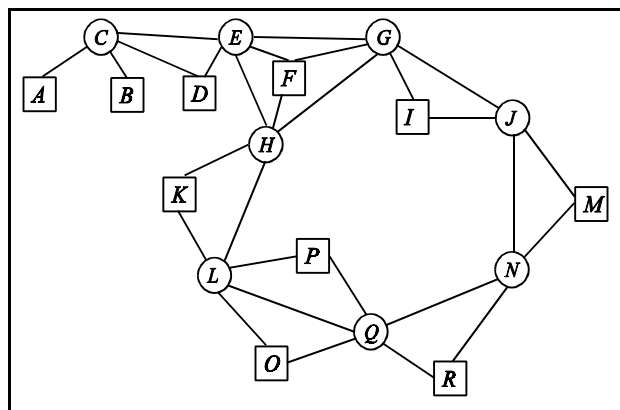


Figure 4: Graph of Entailment Mesh

Gordon Pask’s *entailment mesh* structure is formed by clusters of concepts. Clusters represent the existence of relationships between the concepts they contain. This form of knowledge representation emphasizes that knowledge about each concept entails knowledge about other concepts. The entailment mesh is then a coherent bundle of concepts [Medina-Martins et al, 1993]. A generic mesh of clusters is shown in figure 3.

An alternative way to represent this structure is with a network graph with two different kinds of nodes (figure 4): the *liaison concepts* (circles) and the *terminal concepts* (squares). The first are those

concepts that are included in more than one cluster. This network is basically the structure used by *TalkMine* with keywords and documents. Liaison concepts are the keywords, and terminal concepts the documents. There is, however, one important difference in this structure: documents can become keywords themselves.

<sup>8</sup> A paper delivered at the *Principia Cybernetica Conference* organized by Heylighen, Joslyn, and Turchin at the Free University of Brussels in 1991. Subsequently, the work was developed in different ways by Medina-Martins and Rocha [1992], and Medina-Martins et al [1993, 1994].



entailment meshes associated with specific contexts. The categorization process based on evidence sets previously defined, establishes an extended SelPrune which follows associative chains that are not stored in any one of the individual knowledge contexts (or entailment meshes), bridging together contextually different aspects of knowledge stored in the system. In other words, *TalkMine* as here defined, adds explicit context to Conversation Theory, and as discussed below, a way in for open-ended semiosis.

## 4.2 Open-Ended Semiosis

### 4.2.1 Semantics and Self-Organization

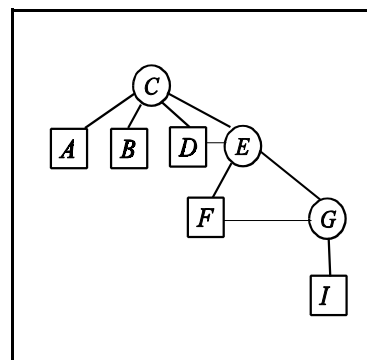
As argued in 3.3.1, the local knowledge contexts of information resources used by *TalkMine* possess an associative semantic metric and high redundancy of semantic information which leads us to see it as a kind of distributed memory, albeit not as in such a strong sense as superposed distribution of memory. In any case, local knowledge contexts do preserve many of the important attributes of distributed memory: a semantic metric constructed from the non-linear integration of many elements (keywords) which are stored across a whole network in a highly redundant manner. The integration of local keyword information into a network-wide semantic metric can be seen as a connectionist process of *self-organization* since it relies on the interaction of many individual components (the documents) to produce a global semantic metric. Any one individual component is redundant in this process. Furthermore, keywords used by the metric are given interpretations by the users of this distributed memory. These interpretations and the keyword associations of the metric (initially constructed from document structure) establish the *semantics* of this distributed memory for the community of users.

### 4.2.2 Pragmatics and Adaptation

As described above, this semantics adapts to users according to their patterns of utilization of the distributed memory, that is, according to the categories they construct in conversation with the system. In other words, the environment of users selects (categorizes, as discussed in 1) the self-organizing semantics of the distributed memory. This way, the semantics of the distributed memory adapts to the evolving expectations of the environment of users which establishes a level of *pragmatics* to the existing semantics, or selected self-organization.

### 4.2.3 Syntax and Categorization

The evidence sets short-term categories used by *TalkMine* together with the question-answering algorithm that uses the evidence set operations mentioned in 2.4, establish a syntax which allows significance to be transmitted across information resources and users. Furthermore, the utilization of this syntax recombines knowledge into short-term categories which are then used to introduce new knowledge into local knowledge contexts through pragmatic adaptation. In other words, this process of categorization offers a mechanism to encode the long-term memory knowledge categories from different information resources into short-term categories. Short-term categories are cast as evidence sets which serve as symbolic carriers. The syntactic rules of evidence sets are then used to manipulate these short-term category representations, recombining



**Figure 8:** Possible *SelPrune* of *C*. *C*, *E*, and *F* are the keywords (liaison concepts) used in the associative chain (the question-answering process of *TalkMine*). The documents (terminal concepts) that best match this particular chain would be retrieved, namely *D* and *F*.

them according to the interests of users in a conversational process. The final categories obtained are then decoded back into the distributed memory banks to produce a list of related documents and to provide an adaptation to the users engaged in conversation, including the introduction of new knowledge into local information resources.

The determination of this encoding from long-term distributed memory into short-term categories to be used in conversation, altered, and the decoded back to long-term memory, is tricky because *everything* in *TalkMine* is at some level a symbol – it is all implemented in a computer after all! But in the biological realm, as discussed in section 1, we also talk of symbolic codes even though everything is implemented at some level as a physical, dynamic, non-symbolic, matter. Therefore, we need to pay extra attention to the functional hierarchies of both of these systems. The long-term distributed memory level keeps knowledge in a networked, associative manner that is quite stable, while the short-term categories in contained sets which are very ephemeral but are ideal for manipulation (unlike distributed memory). The encoding exists as a mediation between these two levels to facilitate transmission (communication) and recombination (creation) of new knowledge, the key goals of a semiotic code as discussed in 1.3<sup>9</sup>.

#### 4.2.4 Linguistic-Based Selected Self-Organization

The evidence set question-answering system of section 6 models the construction of the prototypical effects discussed in section 2. Such “on the hoof” construction of categories triggered by interaction with users, allows several unrelated information resources to be searched simultaneously, temporarily generating categories that are not really stored in any location. The short-term categories bridge together a number of possibly highly unrelated contexts, which in turn creates new associations in the individual information resources that would never occur within their own limited context.

Consider the following example. Two distinct information resources (databases) are going to be searched using the system described above. One database contains the documents (books, articles, etc) of an institution devoted to the study of computational complex adaptive systems (e.g. the library of the *Santa Fe Institute*), and the other the documents of a Philosophy of Biology department. I am interested in the keywords GENETICS and NATURAL SELECTION. If I were to conduct this search a number of times, due to my own interests, the learned category obtained would certainly contain other keywords such as ADAPTIVE COMPUTATION, GENETIC ALGORITHMS, etc. Let me assume that the keyword GENETIC ALGORITHMS does not initially exist in the Philosophy of Biology library. After I conduct this search a number of times, the keyword GENETIC ALGORITHMS is created in this library, even though it does not contain any documents about this topic. However, with my continuing to perform this search over and over again, the concept of GENETIC ALGORITHMS becomes highly associated with GENETICS and NATURAL SELECTION, introducing a new perspective of these keywords. From this point on, users of the Philosophy of Biology library, by entering the keyword GENETIC ALGORITHMS would have their own data retrieval system point them to other information resources such as the library of the *Santa Fe Institute* or/and output documents ranging from “The Origin of Species” to treatises on Neo-Darwinism – at which point they would probably bar me from using their networked database!

---

<sup>9</sup> There are obvious distinctions between biological and linguistic codes as postulated here, namely that in language it is the syntax that is ephemeral in conversation while in biology it is the genes which live on. A discussion of the differences between the different kinds of selected self-organization is left for a future work.



Given a large number of sub-networks comprised of context-specific associations, the categorization system is able to create new categories that are not stored in any one location, changing the long-term memory banks in an open-ended fashion. Open-endedness does not mean that the categorizing system is able to discern all possible details of its user environment, but that it can permutate all the associative information that it constructs in an essentially open-ended manner. Each independent network has the ability to associate new knowledge in its own context (e.g. as more documents are added to the libraries of the prior examples). To this, the categorization scheme adds the ability of open-ended associations built across information resources.

If we regard the *TalkMine*'s learned categories, implemented as evidence sets, as linguistic prototypical categories constructed to integrate the knowledge of a set of information resources with user interests through conversation, then such categories are precisely a mechanism to achieve the linguistic perturbation of long-term distributed memory used to adapt stored knowledge to an evolving environment. In addition, short-term categorization not only adapts an existing structure to its users, but effectively creates new keywords in different, otherwise independent, information resources, solely by virtue of its temporary construction of categories. This way, linguistic categories function as a system of consensual linguistic recombination of distributed memory banks, capable of transferring knowledge across different contexts and thus creating new knowledge. In this way, the full-blown semiosis between DIS and communities of users mediated by *TalkMine* instantiates the linguistic-based selected self-organization described in section 1 and is open-ended as it can adapt to an evolving environment and generate new knowledge given a sufficiently diverse set of information resources and users.

Finally, *TalkMine* is a working recommendation system for DIS following Gordon Pask's goal of an artificial system for driving through knowledge in a new level of adaptive human-machine symbiosis. Readers are encouraged to track the development of this system at <http://www.c3.lanl.gov/~rocha/lww>.

## Acknowledgments

In addition to the intellectual debt I owe Gordon Pask, as it must be obvious throughout my work, I would like to express heartfelt gratitude for all the academic and personal help Gordon was so kind to grant me.

## References

- Balabanović, M. and Y. Shoham [1997]. "Content-based, collaborative recommendation." *Communications of the ACM*. March 1997, Vol. 40, No.3, pp. 66-72.
- Barsalou, L. [1987]. "The instability of graded structure: implications for the nature of concepts." In: *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*. U. Neisser (ed.). Cambridge University Press.
- Chislenko, Alexander [1998]. "Collaborative information filtering and semantic transports." In: . . WWW publication: <http://www.lucifer.com/~sasha/articles/ACF.html>.
- Clark, Andy [1993]. *Associative Engines: Connectionism, Concepts, and Representational Change*. MIT Press.
- Clark, Andy [1997]. *Being There: Putting Brain, Body, and World Together Again*. MIT Press.
- Crutchfield, J.P. and M. Mitchell [1995]. "The evolution of emergent computation." *Proc. National Academy of Sciences, USA, Computer Sciences*. Vol. 92, pp. 10742-10746..
- Dubois, D. and H. Prade [1985]. "A note on measures of specificity for fuzzy sets." *Int. J. of General Systems*. Vol. 10, pp. 279-283.
- Fodor, J. [1981]. *Representations: Philosophical Essays on the Foundations of Cognitive Science*. MIT Press.

- Galvin, F. and S.D. Shore [1991]. "Distance functions and topologies." *The American Mathematical Monthly*. Vol. 98, No. 7, pp. 620-623.
- Hampton, J. [1992]. "Prototype models of concept representation." In: *Categories and Concepts: Theoretical Views and Inductive Data Analysis*. I. Van Mechelen, J. Hampton, R. Michalski, and P. Theuns. Academic Press.
- Henry, C. and L.M. Rocha [1996]. "Language theory: consensual selection of dynamics." In: *Cybernetics and Systems: An International Journal*. . Vol. 27, pp 541-553.
- Johnson, N., S. Rasmussen, C. Joslyn, L. Rocha, S. Smith, and M. Kantor [1998]. "Symbiotic intelligence: self-organizing knowledge on distributed networks, driven by human interaction." In: *6th International Conference on Artificial Life*. C. Adami, et al. (Eds.). MIT Press. In Press.
- Kanerva, P. [1988]. *Sparse Distributed Memory*. MIT Press.
- Kannan, R. and S. Vempala [1999]. "Real-time clustering and ranking of documents on the web." *Yale University*. Unpublished Manuscript..
- Kleinberg, J.M. [1998]. "Authoritative sources in a hyperlinked environment." In: *Proc. of the 9th ACM-SIAM Symposium on Discrete Algorithms*. . pp. 668-677.
- Lakoff, G. [1987]. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. University of Chicago Press.
- Lorenz, K. [1971]. "Knowledge, beliefs and freedom." In: *Hierarchically Organized Systems in Theory and Practice*. P. Weiss (ed.). Hafner.
- Lorenz, K. [1981]. *The Foundations of Ethology*. Springer-Verlag.
- Maturana, H. and F. Varela [1987]. *The Tree of Knowledge: The Biological Roots of Human Understanding*. New Science Library.
- McShea, D.W. [1993]. "Evolutionary change in the morphological complexity of the mammalian vertebral column." *Evolution*. Vol. 47, pp. 730-40.
- McShea, D.W. [1994]. "Mechanisms of large-scale evolutionary trends." *Evolution*. Vol. 48, pp. 1747-63.
- Medin, D.L. and M.M. Schafer [1978]. "A context theory of classification learning." In: *Psychological Review*. . Vol. 35, pp. 207-238.
- Medina-Martins, P., L. Rocha, et al [1994]. "Metalogues: an essay on computers' psychology – from childhood to adulthood." In: *Cybernetics and Systems 94*. R. Trappl (Ed.). World Scientific Press, pp. 565-572.
- Medina-Martins, P. and L. Rocha [1992]. "The in and the out: an evolutionary approach." In: *Cybernetics and Systems 92*. R. Trappl (Ed.). World Scientific Press, pp. 681-689.
- Medina-Martins, P., L. Rocha, et al [1993]. *CYBORGS: A Fuzzy Conversational System*. Final report for the NATO International Program on Learning Systems.
- Morris, C.W. [1946]. *Signs, Language, and Behavior*. G. Braziller Publishers..
- Moukas, A. and P. Maes [1998]. "Amalthaea: an evolving multi-agent information filtering and discovery systems for the WWW." *Autonomous agents and multi-agent systems*. Vol. 1, pp. 59-88.
- Nakamura, K. and S. Iwai [1982]. "A representation of analogical inference by fuzzy sets and its application to information retrieval systems." In: *Fuzzy Information and Decision Processes*. Gupta and Sanchez (Eds.). North-Holland, pp. 373-386.
- Osherson, D. and E. Smith [1981]. "On the adequacy of prototype theory as a theory of concepts." *Cognition*. Vol. 9, pp. 35-58.
- Pask, Gordon [1975]. *Conversation, Cognition, and Learning: A Cybernetic Theory and Methodology*. Elsevier.
- Pask, Gordon [1976]. *Conversation Theory: Applications in Education and Epistemology*. Elsevier.
- Pattee, Howard H. [1995]. "Evolving self-reference: matter, symbols, and semantic closure." *Communication and Cognition - Artificial Intelligence*. Vol. 12, nos. 1-2 (Rocha[1995a]), pp. 9-27.
- Rocha, Luis, M. [1994]. "Cognitive categorization revisited: extending interval valued fuzzy sets as simulation tools concept combination." *Proc. of the 1994 Int. Conference of NAFIPS/IFIS/NASA*. IEEE Press, pp. 400-404.
- Rocha, Luis M. [1996]. "Eigenbehavior and symbols." *Systems Research*. Vol. 13, No. 3, pp. 371-384.
- Rocha, Luis M. [1997a]. *Evidence Sets and Contextual Genetic Algorithms: Exploring Uncertainty, Context and Embodiment in Cognitive and biological Systems*. PhD. Dissertation. State University of New York at Binghamton.
- Rocha, Luis M. [1997b]. "Relative uncertainty and evidence sets: a constructivist framework." *International Journal of General Systems*. Vol. 26, No. 1-2, pp. 35-61.

- Rocha, Luis M. [1998a]. "Selected self-organization and the Semiotics of Evolutionary Systems." In: *Evolutionary Systems: Biological and Epistemological Perspectives on Selection and Self-Organization*. S. Salthe, G. Van de Vijver, and M. Delpo (eds.). Kluwer Academic Publishers, pp. 341-358.
- Rocha, Luis M. [1998b]. "Syntactic Autonomy." In: *Proceedings of the Joint Conference on the Science and Technology of Intelligent Systems (ISIC/CIRA/ISAS 98)*. National Institute of Standards and Technology, Gaithersburg, MD, September 1998. IEEE Press, pp. 706-711.
- Rocha, Luis M. [1998c]. "Where is the progress?." *Cybernetics and Human Knowing*. Vol.5, No. 4, pp. 86-90.
- Rocha, Luis M. [1999a]. "Evidence sets: modeling subjective categories." *International Journal of General Systems*. In Press.
- Rocha, Luis M. [1999b]. "Syntactic autonomy and the Escape from situated action." *Communication and Cognition - Artificial Intelligence*. Submitted.
- Rocha, Luis M. and Cliff Joslyn [1998]. "Models of Embodied, Evolving, Semiosis in Artificial Environments." In: *Proceedings of the Virtual Worlds and Simulation Conference*. C. Landauer and K.L. Bellman (Eds.). The Society for Computer Simulation, pp. 233-238.
- Rocha, Luis M., V. Kreinovich, and K. B. Kearfott [1996]. "Computing uncertainty in interval based sets." In: *Applications of Interval Computations*. V. Kreinovich and K.B. Kearfott (Eds.). Kluwer Academic Publishers, pp. 337-380.
- Rosch, E. [1975]. "Cognitive representations of semantic categories." *J. of Experimental Psychology*. Vol. 104, pp. 27-48.
- Rosch, E. [1978]. "Principles of categorization." In: *Cognition and Categorization*. E. Rosch and B. Lloyd (Eds.). Hillsdale, pp. 27-48.
- Shafer, G. [1976]. *A Mathematical Theory of Evidence*. Princeton University Press.
- Smith, E. and D. Osherson [1984]. "Conceptual combination with prototype concepts." *Cognitive Science*. Vol. 8, pp. 337-361.
- Smith, E.E. and D.L. Medin [1981]. *Categories and Concepts*. Harvard University Press.
- Uexküll, J.V. [1982 [1940]]. "The theory of meaning." *Semiotica*. Vol. 42, no. 1, pp. 25-87.
- Umerez, Jon [1995]. "Semantic Closure: a guiding notion to ground Artificial Life." In: *Advances in Artificial Life*. F. Moran, A. Moreno, J.J. Merelo, and P. Chacon (eds.). Springer-Verlag, pp. 77-94.
- van Gelder, Tim [1991]. "What is the 'D' in 'PDP': a survey of the concept of distribution." In: *Philosophy and Connectionist Theory*. W. Ramsey et al. Lawrence Erlbaum.
- von Foerster, Heinz [1977]. "Objects: tokens for (eigen-)behaviors." In: *Hommage a Jean Piaget: Epistemologie Genetique et Equilibration*. B. Inhelder, R. Gracia, and J. Voneche (Eds.). Delachaux et Niestel. Reprinted in von Foerster [1981], *Observing Systems*, pp. 274-285.
- von Neumann, J. [1966]. *The Theory of Self-Reproducing Automata*. University of Illinois Press.
- Zadeh, Lofti A. [1965]. "Fuzzy Sets." *Information and Control*. Vol. 8, pp. 338-353.